A Content Analysis of Student's Perceptions of Instructors

A Thesis

Presented for the

Master of Science Degree

The University of Tennessee at Chattanooga

Bill C. Cassill

December 1996

To the Graduate Council:

I am submitting a thesis written by Bill C. Cassill entitled "A Content Analysis of Students' Perceptions of Instructors." I have examined the final copy of this thesis and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science with a concentration in Industrial / Organizational Psychology.

[redacted]

Mike Biderman, Chairperson

We have read this thesis
and recommend its acceptance:

[redacted]

[redacted]

Accepted for the Graduate Division:

[redacted]

Director of Graduate Studies

Dedication

I would like to dedicate this thesis to my son, Robert, and my loving wife, Sonja, who steadfastly supported my efforts in this endeavor. I also wish to dedicate this thesis to my mother, Jean Cassill, whose unwavering support made my education possible and to my father-in-law and mother-in-law, Zenovij and Nada Semak, for their gracious hospitality during my final semester in school.

Acknowledgement

I wish to acknowledge the perseverance of my thesis chairperson, Dr. Mike Biderman, who assisted in seeing this project to its conclusion. I also wish to thank the counsel of Dr. Jeff Ryer in the area of structural equations modeling. I wish to thank Dr. Amye Warren for providing me with all of the helpful comments on my thesis, and Dr. Hong for her insights on qualitative data analysis. Finally, I wish to thank Dr. Herbert Marsh, whose work gave me the idea for this thesis.

Abstract

In his recent confirmatory factor analysis of the Instructional Development and Effectiveness Assessment rating instrument (IDEA), Marsh (1994) identified six factors matching those from his Students' Evaluation of Educational Quality (SEEQ) rating instrument. However, four of these factors, Enthusiasm, Interaction, Learning, and Organization, were found to be highly intercorrelated. Due to this, other researchers have questioned whether these four factors are really independent constructs as Marsh asserts. Because of the question of independent constructs, many researchers feel that a greater reliance should be placed on the use of global rating items instead of items designed to measure specific dimensions of instructional effectiveness. Marsh counters with the assertion that responses to global items are nothing more than a weighted average of specific dimensions.

In a parallel line of research, Cadwell and Jenkins (1985) hypothesized that the semantic similarity of individual items was the underlying influence to the robust factor structure found in Marsh's SEEQ and other rating instruments. Their findings suggested that the synonymous wording of items within scales artificially inflates inter-item correlations resulting in an illusory robust factor structure.

This study hypothesized that the use of global open-ended questions in conjunction with the use of the Enthusiasm, Interaction, Learning, and Organization

scales from the IDEA would help disentangle the issues of semantic similarity and of independent constructs. Following a content analysis that categorized responses to the open-ended items into themes that matched the semantic meaning of the four IDEA scales, a correlational analysis revealed that responses to both the closed-ended IDEA scales and the open-ended items possessed fairly good convergent validity effectively disputing the Semantic Item Similarity hypothesis.

Following this, three structural equations models were conducted. The first model demonstrated that a Rater Bias construct representing global response tendencies on the part of student raters accounted for a significant portion of the variance in each of the four scales and offered a possible explanation for the high factor intercorrelations found in Marsh's (1994) study. The second model indicated that the Rater Bias construct also significantly influenced responses to the open-ended items as well. In the final model, a global item was introduced. The global item was found to have significant loadings on the Rater Bias, Learning, and Organization latent variables thereby providing some support to Marsh's assertion that responses to global items are a composite of specific dimensions of teaching effectiveness.

Table of Contents

List of Tables

List of Figures

Introduction

The use of student ratings for instructional evaluation is almost universal in higher education. Student ratings of instructors are used for administrative decisions, instructor feedback, as well as influencing other students' perceptions of course difficulty under a particular instructor. Because of the above uses of rating forms, considerable research has been conducted to develop reliable and valid measures of instructor effectiveness. The two approaches used in the literature to validate rating forms have been the criterion validity approach and the construct validity approach. The criterion validity approach attempts to relate student ratings to other indices of instructional effectiveness, such as ratings by administrators, measures of student achievement, instructor self ratings, etc. (Abrami, d'Apollonia, & Cohen, 1990; Cohen, 1981). The construct validation approach, on the other hand, has attempted to develop psychometrically sound rating instruments designed to measure latent dimensions of teaching effectiveness. However, both approaches have met with conflicting results.

The lack of consistent results in the criterion validity literature is due, in part, to conflicting opinions of the appropriateness of the various criteria listed above (Cashin & Downey, 1992). For example, Marsh (1991b) concluded that high correlations between student ratings and instructor self ratings are evidence of a valid rating instrument while others believe that only student achievement is a suitable criterion (Abrami, d'Apollonia, & Cohen, 1990; Cohen, 1981). The problem is that the use of different criteria often leads to different results. For instance, Cashin and Downey (1992) have found that ratings from the Instructional Development and Effectiveness Assessment System (IDEA), a standardized rating instrument developed at Kansas State University, correlate very highly with students' perceptions of their own learning. It has been shown by other research, however, that actual student learning has only low to moderate correlations with specific rating dimensions and that variance in student learning is mostly accounted for by student ability (Marsh, 1987; as cited by Abrami et al., 1990).

Additionally, it has been suggested that weaknesses found in prior studies have limited their ability to detect the true relationship between student ratings and student achievement (Abrami et al., 1990). Some of these weaknesses include small sample sizes, different operationalizations of rating constructs, failure to control for threats to internal validity, the use of different criteria across different studies, and a failure to represent "real world" instructor characteristics (Abrami et al., 1990).

While one line of research has been to establish the criterion related validity of student ratings, another parallel line of research has focused on establishing the construct validity of such measurement. Prior research has done this mostly through the use of factor analytic techniques (Abrami & d'Apollonia, 1991).

Two perspectives have been formed in the recent literature regarding the construct validity of student ratings. The first perspective views students' perceptions of instructional effectiveness as completely multi-dimensional. One of the most vocal advocates of this perspective has been Marsh at the University of Western Sydney who has conducted considerable research to develop a standardized rating instrument known as the SEEQ (Students' Evaluations of Educational Quality). While research on student ratings dates back to the 1920's (Marsh, 1991a) , the research of Marsh is some of the most recent. Marsh (1987c; as cited by Marsh, 1991a) comments that the similarity of the factor structure across different standardized rating instruments demonstrates that students perceptions of instructional effectiveness is multidimensional. Some of the instruments he has investigated includes Frey's Endeavor Instructional Rating Form (Marsh, 1987c; as cited by Marsh, 1991a), the Michigan State Student Instructional Rating System instrument (Marsh, 1987c; as cited by Marsh, 1991a), and the Instructional Development and Effectiveness Assessment (i.e. IDEA) (Marsh, 1994) developed by Kansas State University. Marsh (1991a) contends that there are nine distinct dimensions students use to rate instructors: Learning, instructor Enthusiasm, Organization of Material, Group Interaction, instructor Rapport, Breadth of Coverage of Material, Exams, Assignments, and Workload. Evidence for the factorial (construct) validity of the SEEQ based on these nine dimensions is substantial (Cadwell and Jenkins, 1985; Marsh, 1991a, 1991b). Due to these findings, Marsh (1991a, 1991b, 1994) contends that each dimension of teaching effectiveness is distinct and unique and that only specific dimensions should be used to rate teaching effectiveness.

The second perspective, which might be called the global construct validity perspective, cautions that complete reliance on ratings of specific dimensions of teaching effectiveness is premature and that more reliance should be placed upon the use of global ratings of instructors due to inconsistent findings in prior research (Abrami et al., 1990; Abrami & d'Apollonia, 1991), the fact that global items tend to correlate more highly with indices of student learning (Cohen, 1988; Cashin, Downey, and Sixbury, 1994), and because a reliance on a factor analytic strategy to demonstrate the construct validity of a rating instrument is inconclusive because of the indeterminacy of factor solutions (Abrami and d'Apollonia, 1991). In spite of the cautions that researchers of the global construct validity perspective give concerning the use of specific dimensions of teaching effectiveness, most researchers sharing this perspective agree that students' perceptions of instructional effectiveness are probably multi-dimensional (Abrami and d'Apollonia, 1991; Cashin & Downey, 1992; Cohen, 1988).

Whereas those who share the global construct validity perspective advocate the use of global rating items, Marsh (1994) believes that responses to global items are nothing more than a weighted average of distinct, independent constructs. The problem with this assertion is that the Enthusiasm, Interaction, Learning, and Organization constructs that Marsh reports as being independent of one another were found to be highly intercorrelated in his recent study of the IDEA (Cashin, Downey, & Sixbury, 1994; Marsh, 1994). Due to this, Cashin et al (1994) imply that some of Marsh's constructs may not be independent of one another. Below is the factor correlation matrix from Marsh's (1994) study of the IDEA rating instrument:

	Learning	Enthusiasm	Organization	Interaction
Learning	---			
Enthusiasm	.85	---		
Organization	.96	.84	---	
Interaction	.80	.80	.76	---

Lisrel 7 partial intercorrelation estimates of latent variables from a confirmatory factor analysis of the IDEA rating instrument (Marsh, 1994, p. 639).

Taken at face value, such results would cast doubt on any researcher's assertion of independent constructs. The nature of these results, however, may have been heavily influenced by way these constructs were measured. Like Marsh's (1994) study, research on student ratings has been based on an almost complete reliance on the use of Likert scaling (Brinko, L'Hommedieu, & Menges, 1990). Systematic rater error, like halo error, that often accompanies the use of these scales is well documented (Alliger & Williams, 1992). The four scales that serve as indicators for these latent constructs listed above are comprised of items that require ratings of specific instructor qualities. A global response tendency on the part of respondents, like general impression bias, may partially account for the high intercorrelations among these four latent constructs.

Another hypothesized confound that may influence responses to closed-ended items that employ Likert response formats is semantic item similarity (Cadwell & Jenkins, 1985). Items that are semantically similar tend to elicit comparable responses due to synonymous item wording. It has been hypothesized that semantic item similarity underlies the robust factor structure of scales developed with factor analytic methods (Cadwell & Jenkins, 1985). Cadwell and Jenkins (1985) investigated this possibility by having students rate written descriptions of instructors using four scales from Marsh's SEEQ rating instrument. After partialling out the effects of treatment, the factor structure of the responses remained relatively unchanged (Cadwell & Jenkins, 1985). This finding led them to conclude that semantic item similarity was a significant contributor to the factor structure. The basis of their conclusion rests on their contention that similar items within a scale received similar responses largely due to general impression of how the items logically covary and not on respondent consideration of specific instructor characteristics or behaviors.

In their reply article to the Cadwell and Jenkins study, Groves and Marsh (1987) were critical of the findings but did concede the possibility that semantic item similarity might influence item intercorrelations within scales. Cadwell and Jenkins (1985) offered corollary evidence to support their findings. First, the use of general impression is typically easier than an exhaustive memory search. Second, most students do not spend much time on any particular item. Third, students probably use a variety of heuristics to reduce cognitive effort in responding to rating items. Findings by Chan (1991), in a separate study, support the idea that respondents typically employ effort reducing heuristics when responding to closed ended questions that offer a series of response options.

The fact that respondents may use effort reducing heuristics has not been something that has been extensively researched or accounted for in the student rating literature. Jenkins (1987) cautions that reliance on a factor analytic strategy to demonstrate construct validity must be tempered with consideration of the cognitive processes of respondents. One of the agreed upon methods to deal with individual differences in perceptions of instructional effectiveness has been the use of class means instead of individual ratings. The assumption behind the use of class means is that individual response styles or biases will average out and that a more accurate account of the constructs underlying instructional effectiveness can be derived (Abrami et al., 1990; Cohen, 1981; Cranton & Smith, 1990). This method would likely reduce the effect of halo error at the individual level, but class means might still be affected by it.

Despite the use of class means, the reliance on the use of closed-ended scales to discern the dimensionality and structure of student ratings places a limitation on their ability to discover independent factors. In other words, the almost total reliance on closed-ended questions in student rating research presents a possible mono-method bias. The use of alternative measures that are not as susceptible to the same response biases found in the use of closed-ended questions may facilitate further investigations of student ratings. If an alternative measure is used in combination with a measure that employs Likert response formats, it might be possible to disentangle the issue of dimensionality from the effects of response bias and from issues of semantic similarity.

One alternative method to the use of scaled rating items in the research on students' perceptions of instructional effectiveness is the use of open-ended items. Responses to open-ended items do not suffer many of the constraints and artifacts found in closed ended questions and, in certain cases, have been found to be more valid indicators of respondents' perceptions and attitudes (Charng & Schaeffer, 1991; Schuman, 1986). The primary reason that most researchers do not rely on the use of open-ended questions in survey research, however, is probably that closed-ended questions are easier to score and analyze. However, content analysis is a means by which responses to open-ended questions can be scored and analyzed.

Content analysis is a qualitative communications research method whereby written or verbal messages can be collapsed into a smaller set of categorical variables or themes. While the method has become more sophisticated through the work of Holsti (1969), Krippendorf (1980), and others, this method of analysis has not been used very frequently in the psychological literature. The use of content analysis has been the almost exclusive domain of sociology-anthropology, general communications, and political science (Holsti, 1969). The reason for this is that much of the research in these disciplines involves either interview techniques or a reliance on written documentation. Successful applications of content analysis include propaganda research during the Second World War (Krippendorf, 1980), the development of the Thematic Apperception Test (Krippendorf, 1980), and the study of successfully aging adults (O'Brian & Conger, 1991).

The method of computer aided content analysis simplifies the categorization and scoring of responses to open-ended items. Computer aided content analysis begins with the construction of a tagging dictionary. A tagging dictionary is a computer text file containing category names and a list of search terms (words or phrases) associated with each. A tagging dictionary is created in accordance with the researcher's assumptions concerning the written material to be analyzed or from a pre-existing dictionary data base. The content analysis software then analyzes the document data file (in this case, containing the typewritten responses to the open-ended items) by comparing words and phrases in the document with the list of search terms. Those that match are tagged by having the corresponding category label inserted next to them in the output file. By this method, the researcher can code responses into data for further statistical analysis.

The Present Study

The software used to conduct the content analysis in this study is called Verbatim Pro and was developed by Dr. Mark Neale, a Professor of Journalism at the University of Tennessee at Knoxville. A more detailed description of the software is included in Appendix B.

The first goal of this study was to discern the extent to which individual differences in global response tendencies influence responses to a set of closed-ended questions with a Likert response format. This was done using the Enthusiasm, Interaction, Learning, and Organization scales from the IDEA (the full questionnaire used in this study can be found in Appendix A). The primary reason for using these four scales is that the four constructs they were intended to represent were found to be highly intercorrelated in Marsh's (1994) study. These four scales are defined as follows:

Learning/Value: the instructor stimulates interest and intellectual effort / uses examples,
Enthusiasm: instructor expressiveness (body language and expressiveness of voice),
Organization: how well a course "hangs together"; the logical order of class lectures and the presentation of class materials; an emphasis on main points and course objectives,
Interaction: how much the instructor encourages student comments, questions, and discussion; explains criticisms of academic performance.

The effects of individual differences on response tendencies can be estimated for each of the four scales through the use of a structural equations model that controls for instructor effects on the four latent variables each of the scales is supposed to represent. It was hypothesized that a Rater Bias construct representing individual differences in response tendencies would partially explain the unusually high factor correlations found between the Enthusiasm, Interaction, Learning, and Organization constructs.

A second goal was to determine the extent to which responses to the closed-ended items are influenced by semantic item similarity. If semantic item similarity is responsible for a large portion of the common scale variance within each of the four scales, there should be a lack of convergent validity between the four scales and responses to the open-ended items used in this study. The convergent validity of the two measures was assessed in two ways. First, simple correlations were computed between the four IDEA scales and a set of variables derived from the open-ended items that match the semantic meaning of the of the four IDEA scales. Second, the same open-ended variables were also included in the structural equations model and allowed to load onto the latent variables that the closed-ended scales theoretically represent. If the Enthusiasm, Interaction, Learning, and Organization dimensions of instructional effectiveness do exist, then some mention of them should be made in the open-ended items. In addition, the open-ended items should not be influenced by semantic item similarity since they lack the stimuli (i.e. identical or synonymous wording) thought to be the cause of this phenomenon.

Two open-ended questions were used: one asking for positive aspects of the instructor while the other asked for negative aspects. The reason for the use of two open-ended questions was that category dimensions derived from the responses can be scored along a bipolar continuum. Frequency counts for each subject's response to the positive item along a particular dimension were given a positive score, while frequency counts along the same dimension were given a negative score for the negative item. In this way, frequency counts from the two questions could be summed for a particular dimension, thereby producing a quantifiable score.

The third goal of this study was to determine whether responses to global items are a composite of specific teaching dimensions. Marsh (1994) and Cashin and Downey (1992) tested this assertion through the use of regression analysis. However, regression analysis is not a suitable means of estimating and partialling out the effects of measurement error. A more appropriate analysis is the structural equations model outlined in this study. It should provide a better estimate of the influence that Rater Bias and the other four latent variables of interest have on responses to the global item included in this study.

Method

Subjects

Two hundred twenty subjects from seven different classes at the University of Tennessee at Chattanooga were surveyed. No attempt was made to separate subjects into groups of gender, age, or ethnicity since these characteristics generally have not been found to be biasing factors in the student rating literature (Abrami et al., 1990).

Procedure

This study was a survey research design. The study was performed during in class sessions whose instructors agreed to participate. After introducing the researcher, the instructor left the room. Following this, the researcher explained to the class that this study was investigating student ratings and that students' individual responses would remain confidential. Also, they were told that participation was not mandatory and that they could stop participating at any time. As a final instruction, they were asked to fill out the questionnaire as conscientiously as possible.

Within each class, two alternate forms of the questionnaire were used. One half of the students received the global item followed by the scaled items and the open-ended questions, in that order. The other half received the same set of questions except that the presentation order of the open-ended items and the scaled items was reversed. The reason for the two forms of the survey questionnaire was to assess possible carryover effects due to the ordering of items.

A content analysis was performed on responses to the open-ended items using the Verbatim Pro software package (a more complete description of this method and the software package are provided in Appendix B). Responses were placed into category variables that matched the semantic content of the four scales used in this study. The operationalizations of these category variables appears below:

organization: tying together relevant information (including book / class lecture) and examples so that topics are understood; using class time effectively including punctuality and not going over allotted class time; the ability to break down abstract or complex topics into more simple terms; providing additional study aids that help to organize information including outlines, study sheets, etc.,

interaction: taking an interest in students' well being; friendliness; the ability to relate to students; involving the class in discussion and Q&A sessions,

enthusiasm: having lively and animated presentation skills,

learning: stimulating student interest during lecture through the use of examples, demonstrations, experiments, studies, and other audio-visual aids; how up to date are studies and other learning aids.

In categorizing responses into each of the above dimensions, an attempt was made to represent the four closed ended scales as accurately as possible. However, some phrases were included in each which, in the subjective judgement of the author reflected the intent of the scale. The largest discrepency was with regard to the Interaction dimension. In categorizing responses to this dimension, we tried to include responses which reflected the items in the IDEA scale, but we also included words and phrases from the "Rapport" and "Interaction" dimensions as defined by Cohen (1981). This was done since both seemed to co-occur in responses to the open-ended items. Several subjects implied that their instructor's friendly attitude toward the class made them feel more free to ask questions and engage in class discussions.

Statistical Analysis

A multivariate T-test was performed across scale items and category variables to assess for possible carryover effects due to form type. In addition, correlations were computed between scale scores and category variables to assess possible differences in common variance due to form type. Reliability measures were then computed for each scale to assess internal consistency. Next, simple correlation co-efficients were computed between the global item, the composite scale scores (items had been averaged), and the category variables derived from the content analysis. Finally, the structural equations models were tested using Amos developed by James Arbuckle (1997).

Results and Discussion

The percentage of respondents making reference to each of the category variables from the open-ended items differed substantially. Organization was mentioned by 57% of the respondents followed by Learning (40%), Interaction (33%), and Enthusiasm (26%) (means and standard deviation of scale scores and category variables are reported in Table 11, p. 42). A preliminary interpretation was that Organization and Learning were, across respondents, considered to be more important to effective instruction than Interaction and Enthusiasm.

The results of the multivariate T-test indicated that there were no significant differences in category and item means associated with the two different ordering of items (Likert items vs. open-ended questions first) (Hotelling's T=.04009, p.=.996). The correlations computed for the two groups given the two form types for the global item, scale scores, and category variables (in Appendix E) revealed that there were somewhat higher correlations between variables for the closed-ended first form type versus the open-ended first form type. However, the pattern and significance of correlations between variables were consistent in both matrices.

The next analysis to be performed was the reliability analysis of the four scales taken from the IDEA. The scale with the highest reliability co-efficient was the Organization scale (α =.8636). This was followed by the Interaction scale (α =.8507), the

Learning scale (α=.7820), and the Enthusiasm scale (α=.7181). The scales for Interaction and Organization have acceptable alpha levels. While the Enthusiasm and Learning scales both have reliability co-efficients less than desired, the Learning scale is closer to an acceptable level than the Enthusiasm scale. A possible reason for the differences in the reliability ratings would be the number of items in each scale. The Interaction and Organization scales were comprised of five and four items, respectively. The Enthusiasm and Learning scales, however, were comprised of only three items each.

Correlations were computed between the averaged scale scores and the dimensions derived from the content analysis (see Table 1, p. 33). In and of themselves, the high intercorrelations between scale scores might suggest a unidimensional rating construct. However, the correlations between the open-ended category variables are near zero which indicates that the constructs may be independent. The moderately high intercorrelations between the closed-ended scale scores suggest the influence of the hypothesized rater bias. The results of the analysis indicate fairly good convergent validity between closed-ended scale scores and the open-ended category variables: each scale correlates significantly with its open-ended counterpart. All of the open-ended category variables, except enthusiasm, have the highest correlation with the closed-ended counterparts. The global item is significantly correlated with all of the scale scores and category variables. This further suggests that global items may represent a composite score of several independent dimensions. However, its higher correlations with the scale scores may also be attributable to the rater bias construct.

The initial structural equations model (depicted on p. 34), reflecting the hypothesises that responses to open-ended items were not influenced by rater bias, contained the scale scores from the IDEA, open-ended scales, the latent variables representing the Enthusiasm, Interaction, Learning, and Organization constructs, a latent variable representing the Rater Bias construct, and six effects coded instructor variables. Following standard procedures for coding multiple groups, six group coding variables were required to code seven instructors. Since the amount of variance accounted for in the dependent (endogenous) variables is the same no matter which coding scheme is used (McClendon, 1994), effects coding was used so that each variable would represent the uniqueness of an instructor relative to the group average. Paths from the latent constructs to the closed-ended scales were fixed to unity since they tended to possess more variability. The path from the rater bias construct to the Enthusiasm scale was also fixed to unity. There were no paths from the rater bias construct to the open-ended category variables in this model.

The fit indices indicated a reasonably good fit of the model (model X^2=89.695 with df=42, (p.=.0000), CFI=.963, GFI=.941, TLI=.920, NFI=.935). Given that the Comparative Fit Index (i.e. CFI) is the fit index of choice (Bentler, 1990b as cited by Byrne, 1994), more weight was placed on it for the fit of the model than any of the other fit indices. However one result of the model was that the enthusiasm open-ended category variable did not significantly load onto the Enthusiasm construct ($t < 1.96$).

To test their independence from the rater bias construct, the open-ended measures were allowed to load onto the rater bias construct as well (see p. 36). Surprisingly, all of the loadings for the open-ended measures on the rater bias construct were significant. This indicated that rater bias was also influencing responses to the open-ended measures as well, though to a much smaller degree. Allowing the open-ended measures to load on the rater bias construct also significantly increased the fit of the model (incremental X^2=40.13 with df=4, p.=.000) (model X^2=49.565 with df=38, (p.=.0991), GFI=.964, CFI=.991, NFI=.964, TLI=979). The significance of the hierarchical chi-square suggested that the open-ended items were influenced by rater bias. In addition, the model chi-square is also non-significant (p.=.099).

The results of the second structural equations model indicated that the rater bias construct had a significant influence on responses to each of the four closed-ended scales. The loadings for the Interaction and Enthusiasm scales were particularly high (.700 and .724, respectively). The loadings for the Learning and Organization scales were somewhat lower but were still significant (.567 and .583, respectively). The loadings suggested that the within scale variance accounted for by the rater bias construct varied from 32% to 52% dependent upon the individual scale.

The loadings of the open-ended category dimensions on the rater bias construct were significant but uniformly much lower than those of the closed-ended scales. Loadings for the category variables ranged from .188 for open-ended Enthusiasm to .296 for open-ended Organization. Although responses to the open-ended items were influenced by rater bias, they were influenced to a much smaller degree than the closed-ended scales.

The second structural equations model also demonstrated the convergent validity of the two measures. All of the open-ended variables (except for open-ended Enthusiasm) had significant loadings on their respective latent variables. Standardized loadings on the respective latent variables for the Organization, Learning, Interaction, and Enthusiasm open-ended were .426, .406, .286, and .108. The loadings for the closed-ended scales were comparatively higher, however. The loadings for the Organization, Learning, Interaction, and Enthusiasm scales were .772, .678, .714, and .689, respectively. The reason for the differences loadings may be the low reliability associated with the use of open-ended items.

Finally, the global item was added to the second model (see p. 38). This model also had a good fit (model X^2=58.679 (p.=.118) with df=47, CFI=.992, GFI=.961, NFI=.963, TLI=.983). The loadings for the global item on the Enthusiasm, Interaction, Learning, Organization, and Rater Bias latent variables were .023,.046, .387, .333, and .499, respectively. However, the only significant loadings were on the Learning, Organization, and Rater Bias constructs. While Learning seemed to have a marginally greater influence on responses to the global item than Organization, the Rater Bias construct had the largest influence of the three.

Finally, there were differences in perceptions of teaching styles across the six instructor variables. While instructor #1 was perceived to be significantly higher than the group average for instructors in Enthusiasm, instructor #2 did not markedly differ from the group average on any of the four dimensions. Instructor #3 was perceived to be higher than the group average in Enthusiasm, Interaction, and Learning but not in

Organization. Instructor #4 was perceived to be higher than average on all four dimensions of teaching effectiveness, while instructor #5 was perceived to be high in Learning and Organization but lower than average in Interaction. Instructor #6, though, was perceived to be significantly lower than the group average in Learning. These results indicate that students differ in their perceptions of teaching effectiveness across different instructors. These results also indicate that a portion of the relationship between these four dimensions of teaching effectiveness can be explained by differences between instructors since some instructors were perceived to be higher (or lower) than others in levels of the four constructs.

The first goal of this study was to determine the extent to which individual differences in response styles influenced responses to a set of closed-ended items that used a Likert response format. Individual differences in global response tendencies represented by a Rater Bias construct were shown to have a significant influence on responses to the four scales used in this study. However, it was also demonstrated that individual differences influenced responses to the open-ended items as well, though to a much smaller degree. The high intercorrelations Marsh (1994) found between the Enthusiasm, Interaction, Learning, and Organization constructs may be partially explained by the Rater Bias construct introduced in this study. Given that his confirmatory factor analysis did not include a rater bias factor, much of the common variance among the scaled items accounted for by rater bias would have been redistributed to the factor intercorrelations.

In this model, rater bias could represent the tendency to give uniform responses across both the closed-ended scale items and the open-ended questions. The correlations between the measures and the positive loadings on the Rater Bias construct indicate that an individual who tended to rate an instructor high on the four scales would also tend to include more positive than negative responses in the open-ended items. The same would be true of someone who tended to rate an instructor uniformly negative on the four scales in that they would tend to include more negative than positive comments in the open-ended questions also. While rater bias seems to heavily influence responses to closed-ended questions using a Likert response format, its effect seems much smaller on responses to open-ended items.

The theory of semantic item similarity was not supported. Theoretically, the open-ended items are not influenced by semantic item similarity since no words were presented as stimuli. Yet there are significant correlations between each scale score and its open-ended category dimension which indicates that both measures are assessing similar constructs. In addition, the significant loadings for the two types of measures on the Interaction, Learning, and Organization factors further supports this finding. While semantic item similarity may have a small influence on the way students respond to closed-ended questions on instructor rating forms, the effect of semantic item similarity seems to be nowhere near as large as suggested by Cadwell and Jenkins (1985).

Marsh's assertion that responses to global items are a weighted average of specific latent dimensions is supported due to the global item's significant loadings on the Learning and Organization constructs. However, this study showed that rater bias heavily influenced responses to this item as well. Since only four of Marsh's nine hypothesized dimensions were included in this study, however, exactly which specific dimensions influence responses to global items remains unclear.

As hypothesized by Marsh and others, it seems that students' perceptions of instructional effectiveness are multidimensional. The three structural equations models in this study adds additional support to this assertion. However, stating what these dimensions are should be tempered with the knowledge that what is being measured is not directly observable. One of the weakest links in research of this kind is the labeling and operationalization of latent constructs. Even though a supposed latent variable exists does not mean that we understand it or even have it labeled correctly (Cliff, 1983; as cited by Loehlin, 1992).

This study demonstrates the need to use alternative measures to assess within subject perceptions of instructional effectiveness. Although the effects of rater bias are well documented in scales developed using internal consistency measures (Alliger & Williams, 1992), the student rating literature has placed an almost total reliance on items that employ Likert scaling procedures. The ease with which these types of items are constructed and scored belies the hidden danger in their use. While the closed-ended items seemed to be superior measures of the latent variables in the present study, rater bias accounted for a very large percentage of the variance within each of the scale scores. Further studies on student ratings should attempt to account for rater bias. However, the intriguing finding that rater bias also influenced responses to the open-ended items was surprising. This result should also serve as a caution for researchers who rely largely on qualitative data analysis techniques.

This study has successfully demonstrated how a qualitative data analysis technique can be incorporated into a study using quantitative measures. As shown by the correlational analysis, the use of open-ended items to supplement closed-ended items can be a powerful means to demonstrate the construct validity of a questionnaire. Furthermore, it can also help to put responses to closed-ended items into a more meaningful context.

In conclusion, it is hoped that the results of this study will serve as a consideration to researchers and practitioners who use closed-ended questionnaires. Systematic error in the form of rater bias potentially can result in misleading conclusions on the part of the researcher who develops a questionnaire to measure multiple constructs.

References

Abrami, P.C. and d'Apollonia, S. (1991). Multidimensional students'evaluations of teaching effectiveness--Generalizability of "N=1" research: comment on Marsh (1991). Journal of Educational Psychology, 83(3), 411-415.

Abrami, P. C., d'Apollonia, S., and Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. Journal of Educational Psychology, 82(2), 219-231.

Alliger, G. M. and Williams, K. J. (1992). Relating the internal consistency of scales to rater response tendencies. Educational and Psychological Measurement, 52, 337-343.

Arbuckle, J. (1997) Amos users' Guide Version 3.6. Chicago, IL: SmallWaters Corp.

Brinko, K. T., L'Hommedieu, R., and Menges, R. J. (1990). Methodological explanations for the modest effects of feedback from student ratings. Journal of Educational Psychology, 82(2), 232-241.

Byrne, B. M. (1994). Structural Equations Modeling with EQS and EQS / Windows. (p. 55). Thousand Oaks, CA: Sage Publications, Inc.

Cadwell, J. and Jenkins, J. (1985). Effects of the semantic similarity of items on student ratings of instructors. Journal of Educational Psychology, 77(4), 383-393.

Cashin, W. E. and Downey, R. G. (1992). Using global student rating items for summative evaluation. Journal of Educational Psychology, 84(4), 563-572.

Cashin, W. E., Downey, R. G., and Sixbury, G. R. (1994). Global and specific ratings of teaching effectiveness and their relation to course objectives: reply to Marsh (1994). Journal of Educational Psychology, 86(4), 649-657.

Chan, J. C. (1991). Response order effects in Likert-type scales. Educational and Psychological Measurement, 51, 531-539.

Charng, H. W. and Schaeffer N. C. (1991). Two experiments in simplifying response categories: Intensity ratings and behavioral frequencies. Sociological Perspectives, 32(2), 165-182.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 51(3), 281-309.

Cohen, P. A. (1988, August). The Student's Role in Evaluating Teaching. Paper presented at the Continuing Education Workshop for the American Psychological Association. Atlanta, GA.

Cranton, P. and Smith, R. A. (1990). Reconsidering the unit of analysis: a model of sudent ratings of instruction. Journal of Educational Psychology, 82(2), 207-212.

Groves, M. A. and Marsh, H. W. (1987). Students' evaluations of teaching effectiveness and implicit theories: A critique of Cadwell and Jenkins (1985). Journal of Educational Psychology, 79(4), 483-489.

Holsti, O. R. (1969). Content Analysis for the Social Sciences and Humanities. (p. 21). Reading, MA: Addison-Wesley Publishing Company, Inc.

Jenkins, J. (1987). Implicit theories and semantic similarities: Reply to Marsh and Groves (1987). Journal of Educational Psychology, 79(4), 490-493.

Krippendorf, K. (1980). Content Analysis: An Introduction to Its Methodology. (pp. 16, 18, 110). Beverly Hills: Sage Publications, Inc.

Loehlin, J. C. (1992). Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis. (pp. 227, 229). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H. W. (1991a). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures. Journal of Educational Psychology, 83(2), 285-296.

Marsh, H. W. (1991b). A multidimensional perspective on students' evaluations of teaching effectiveness: Reply to Abrami and d'Apollonia (1991). Journal of Educational Psychology, 83(3), 416-421.

Marsh, H. W. (1994). Weighting for the right criteria in the Instructional Development and Effectiveness Assessment (IDEA) System: Global and specific ratings of teaching effectiveness and their relation to course objectives. Journal of Educational Psychology, 86(4), 631-648.

McClendon, M. J. (1994). Multiple Regression and Causal Analysis. (p. 223). Itasca, IL: F. E. Peacock Publishers, Inc.

O'Brian, S. J. and Conger P. R. (1991). No time to look back: Approaching the finish line of life's course. International Journal of Aging and Human Development, 33(1), 75-87.

Schuman, H. (1986). Ordinary questions, survey questions, and policy questions. Public Opinion Quarterly, 50(3), 432-442.

Appendix A

With regard to the instructor, please rate how often your instructor did each of the following using the rating format below:

1-hardly ever, 2-occasionally, 3-sometimes, 4-frequently, 5-almost always

1. Promoted teacher-student discussions (as opposed to mere responses to questions)
dimension (interaction)
2. Found ways to help students answer their own questions.
dimension (interaction)
3. Encouraged students to express themselves free and openly.
dimension (interaction)
4. Seemed enthusiastic about the subject matter.
dimension (enthusiasm)
7. Spoke with expressiveness and variety in tone of voice.
dimension (enthusiasm)
9. Made presentations that are dull and dry.
dimension (enthusiasm) / reversed scored
10. Made it clear how each topic fits into the course.
dimension (organization)
11. Explained the reasons for criticisms of students' academic performance.
dimension (interaction)
13. Encouraged student comments even when they turn out to be incorrect or irrelevant.
dimension (interaction)
14. Summarized material in a way which aids retention.
dimension (organization)
15. Stimulated students to intellectual effort beyond that required by most courses.
dimension (learning)
16. Clearly stated the objectives of the course.
dimension (organization)
17. Explained the course material clearly, and explanations are to the point.
dimension (organization)
18. Related course material to real life situations.
dimension (learning)
20. Introduced stimulating ideas about the subject.
dimension (learning)

(instructions have been modified to match the conditions of this study)

These items were taken from Cashin and Downey (1992, p. 571). They comprise the instructor scale of the IDEA rating instrument. Although the instructions have been modified, they are essentially the same as the original. The dimensions listed below each item were taken from Marsh's (1994) study of the IDEA instrument. This is a standardized rating instrument. Marsh's own study used data obtained from 29,543 university classes. Some of the items have been deleted since they did not correspond to any of the four scales used in this study. Negatively worded items are reversed scored (Marsh, 1994).

(Global item)

Overall, how would you rate this instructor?

terrible poor average good excellent

This is the global instructor item. I wrote this one.

(Open-ended items)

1. In what aspects do you think your instructor a good teacher (i.e. what specific things does he or she do well)?

2. In what aspects do you think your instructor needs improvement (i.e. what specific things does he or she not do well)?

These are the two open-ended items. The items are global enough to be relatively context free but still allow responses to be categorized as either positive or negative instructor qualities.

Appendix B

Verbatim Pro was the software application to perform the content analysis in this study. The software allows a variety of units of analysis (i.e. paragraph, sentence, etc.) as well as a variety of scoring methods. For purposes of this study, frequency counts were used as a measure of the degree of affect associated with each dimension.

The data file consisted of type written responses from the questionnaires. Actually, two data files were used: one for each open-ended item. Below is a portion of one of the data files. Notice the "#" sign followed by a coded heading at the top of each response. The pound sign indicates to Verbatim Pro that the text below it is from a new respondent.

#1Instructor1

Very enthusiastic about the subject. Lectures were to the point and easy to follow. Very knowledgeable about subject. She is one of the best instructors I have had at UTC.

#2Instructor1

She gives approaches the topics holistically; she attempts to integrate all possible aspects of a given topic.

#3Instructor1

She presents the book well. When you read the text, you can relate what you have read with the notes from class. She answers questions clearly and tries not to confuse you. She is always willing to help with any problem you may have and always comes to class with a very positive, uplifting attitude. Overall, I feel she does a great job. Her tests are fair and her notes are thorough.

#4Instructor1

She gives great lectures that are well prepared and delivered. Sometimes the material was dry, but delivered well. And the material would have been worse given by other professors.

Before the software application can process the raw text file, it must first be formatted.

Below is an example of a formatted file:

```
#1Instructor1
Very enthusiastic about the subject .
Lectures were to the point and easy to follow .
Very knowledgeable about subject .
She is one of the best instructors I have had at UTC .

#2Instructor1
She gives approaches the topics holistically ; she attempts to integrate all possible aspects of a
given topic .

#3Instructor1
She presents the book well .
When you read the text , you can relate what you have read with the notes from class .
She answers questions clearly and tries not to confuse you .
She is always willing to help with any problem you may have and always comes to class with a
very positive , uplifting attitude .
Overall , I feel she does a great job .
Her tests are fair and her notes are thorough .
```

As stated in the introduction, a search dictionary is also necessary. Basically, this is simply a list of words or phrases the software application uses for tagging and coding purposes. Below is a portion of the tagging dictionary used for the enthusiasm dimension:

```
>>ENTHSM<<
passion
exuberance
confidence
added life
forceful
upbeat
want to be here
not just a job
bland
pep
```

An essential part of the analysis is called a KWIC (key word in context). This is an iterative process whereby tagged words and phrases that have no relation to the construct of interest are deleted from the analysis. Verbatim Pro allows the user to do this by placing brackets ([ ]) around the erroneous terms in the raw data file and reformatting it. The bracketed material is deleted from the formatted file. Below is output associated with the key word in context:

```
#12Instructor1
She is very knowledgeable about the subject .
Probably has done much work in this area , obviously .
She seems to be very > ENTHSM excite< d about the subject .


#21Instructor1
She seems to truly > ENTHSM enjoy<  the course , which causes students to be a little more >
ENTHSM enthu< siastic about learning the subject matter .


#25Instructor1
Does prepare the class through emphasis in lectures .
Displays a genuine interest in the course .
Will work individually with students in small groups away from regular class time .
Covers material completely and with > ENTHSM enthu< siasm .
Seems to have a thorough knowledge of the field .


#28Instructor1
Very > ENTHSM enthu< siastic about the subject .
```

Once a reasonable fit is has been found, Verbatim Pro allows the user to save the frequency counts for each case to an outfile that can be used by a variety of statistical applications including SPSS, SAS, etc.

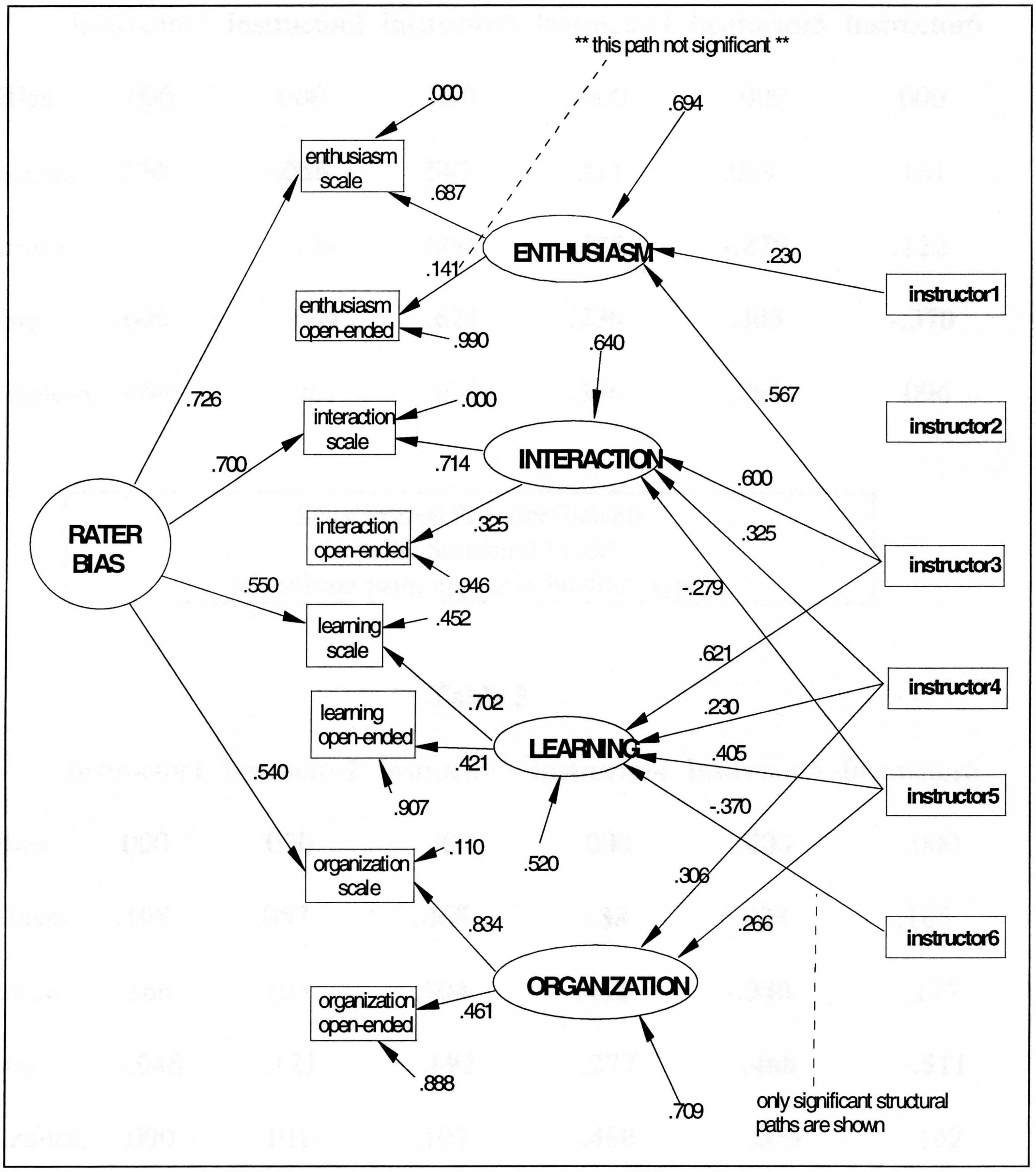
For those interested in using this software, it is free to those who wish to use it for academic purposes. It can be downloaded over the Internet.

Table 1

	global item	enthusiasm scale	interactio n scale	learning scale	organization scale	enthusiasm open-ended	interaction open-ended	learning open-ended	organization open-ended
global item	1.00								
enthusiasm scale	.64**	1.00							
interaction scale	.62**	.78**	1.00						
learning scale	.72**	.64**	.61**	1.00					
organization scale	.71**	.63**	.61**	.69**	1.00				
enthusiasm open-ended	.21**	.21**	.11	.17*	.15*	1.00			
interaction open-ended	.30**	.29**	.38**	.23**	.20**	-.05	1.00		
learning open-ended	.35**	.30**	.24**	.40**	.28**	-.01	.06	1.00	
organization open-ended	.40**	.28**	.28**	.35**	.50**	.08	.09	.11	1.00

* - p. < .05 ** - p. < .01

Correlation Matrix of Observed Variables



The Initial Model

Table 2

	Instructor1	Instructor2	Instructor3	Instructor4	Instructor5	Instructor6
Rater Bias	.000	.000	.000	.000	.000	.000
Enthusiasm	**.230**	-.056	**.567**	.211	.029	-.161
Interaction	.137	-.134	**.600**	**.325**	**-.279**	.122
Learning	-.040	.089	**.621**	**.230**	**.405**	**-.370**
Organization	.064	.061	.079	**.306**	**.266**	.096

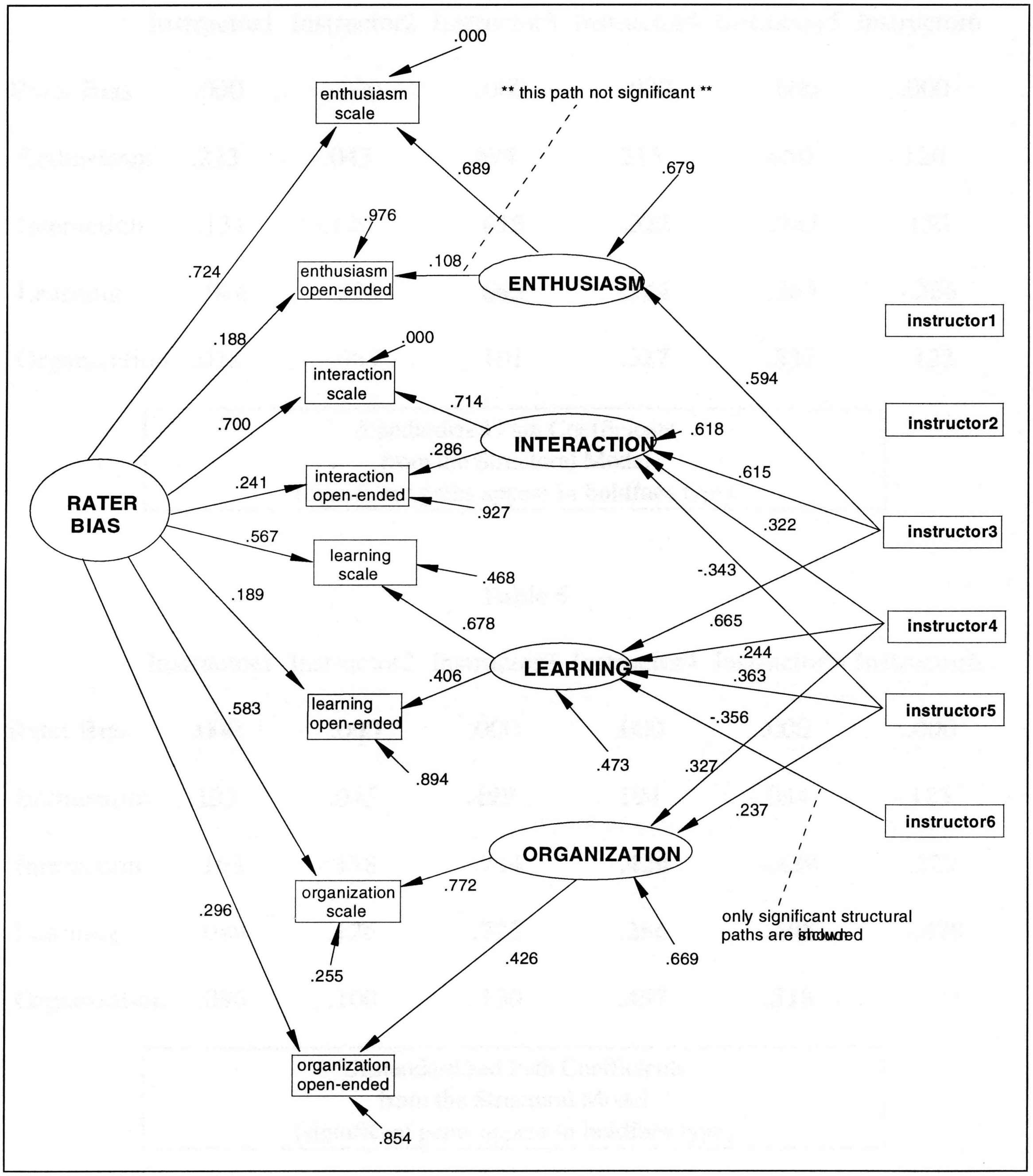
Standardized Path Coefficients
from the Structural Model
(significant paths appear in **boldface** type)

Table 3

	Instructor1	Instructor2	Instructor3	Instructor4	Instructor5	Instructor6
Rater Bias	.000	.000	.000	.000	.000	.000
Enthusiasm	**.195**	-.057	**.468**	.188	.024	-.165
Interaction	.166	-.193	**.704**	**.412**	**-.340**	.177
Learning	-.046	.121	**.692**	**.277**	**.468**	**-.511**
Organization	.090	.101	.107	**.450**	**.375**	.162

Unstandardized Path Coefficients
from the Structural Model
(significant paths appear in **boldface** type)

Correlation estimates for the instructor variables are included on the last page of Appendix D



The Second Model
with category variables loading onto
the Rater Bias Construct

Table 4

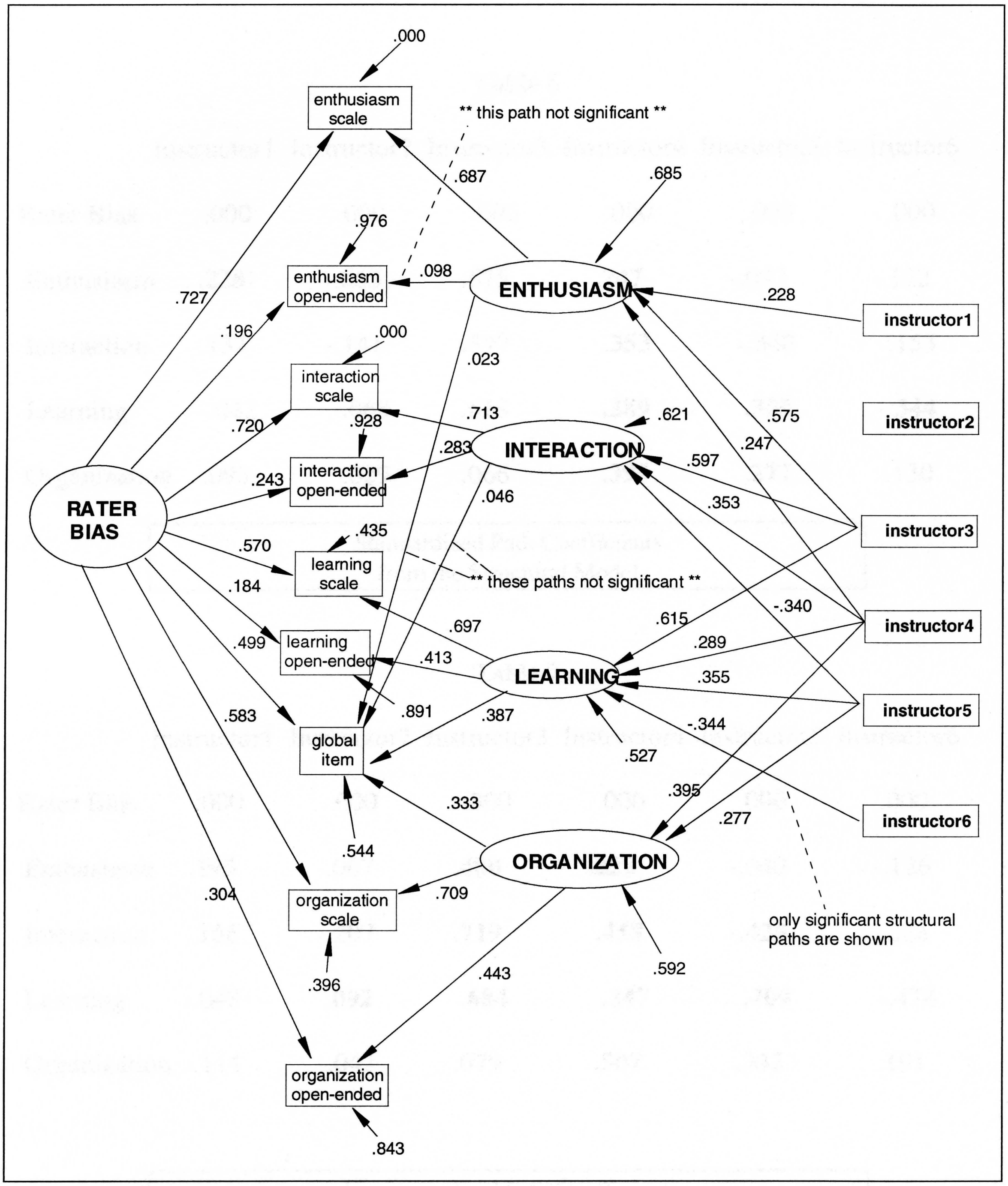
	Instructor1	Instructor2	Instructor3	Instructor4	Instructor5	Instructor6
Rater Bias	.000	.000	.000	.000	.000	.000
Enthusiasm	.223	-.043	**.594**	.215	-.050	-.120
Interaction	.131	-.120	**.615**	**.322**	**-.343**	.153
Learning	-.044	.095	**.665**	**.244**	**.363**	**-.356**
Organization	.072	.063	.101	**.327**	**.237**	.123

Standardized Path Coefficients
from the Structural Model
(significant paths appear in **boldface** type)

Table 5

	Instructor1	Instructor2	Instructor3	Instructor4	Instructor5	Instructor6
Rater Bias	.000	.000	.000	.000	.000	.000
Enthusiasm	.193	-.045	**.499**	.194	-.044	-.125
Interaction	.163	-.178	**.744**	**.420**	**-.429**	.229
Learning	-.049	.126	**.722**	**.286**	**.408**	**-.479**
Organization	.096	.100	.130	**.457**	**.318**	.197

Unstandardized Path Coefficients
from the Structural Model
(significant paths appear in **boldface** type)



The Final Model
With the Global Item Included

Table 6

	Instructor1	Instructor2	Instructor3	Instructor4	Instructor5	Instructor6
Rater Bias	.000	.000	.000	.000	.000	.000
Enthusiasm	**.228**	-.065	**.575**	**.247**	-.047	-.122
Interaction	.134	-.141	**.597**	**.353**	**-.340**	.153
Learning	-.042	.068	**.615**	**.289**	**.355**	**-.344**
Organization	.093	.029	.066	**.395**	**.277**	.130

Standardized Path Coefficients
from the Structural Model

Table 7

	Instructor1	Instructor2	Instructor3	Instructor4	Instructor5	Instructor6
Rater Bias	.000	.000	.000	.000	.000	.000
Enthusiasm	**.195**	-.067	**.480**	**.222**	-.040	-.126
Interaction	.166	-.207	**.719**	**.458**	**-.425**	.228
Learning	-.048	**.092**	**.684**	**.347**	**.409**	**-.474**
Organization	.114	.042	.079	**.507**	**.342**	.191

Unstandardized Path Coefficients
from the Structural Model

Table 8

	instructor1	instructor2	instructor3	instructor4	instructor5	instructor6
instructor1	1.000					
instructor2	.579	1.000				
instructor3	.471	.569	1.000			
instructor4	.510	.599	.497	1.000		
instructor5	.489	.582	.475	.513	1.000	
instructor6	.587	.664	.577	.606	.590	1.000

Correlation matrix of instructor variables

(only reported once since they were the same for all models)

Table 9

	global item	enthus scale	interact scale	learn scale	organ scale	enthus OE	interact OE	learn OE	organ OE
global item	1.0000								
enthusiasm scale	.5664**	1.0000							
interaction scale	.5716**	.7783**	1.0000						
learning scale	.7205**	.6448**	.6026**	1.0000					
organization scale	.7033**	.6399**	.5725**	.6532**	1.0000				
enthusiasm OE	.1844	.2075*	.1581	.2203*	.1585	1.0000			
interaction OE	.2801**	.3222**	.4334**	.2381*	.1432	-.0781	1.0000		
learning OE	.3723**	.3235**	.1755	.4775**	.2858**	.0832	-.0115	1.0000	
organization OE	.3791**	.3299**	.2979**	.3247**	.5141**	.0494	.1116	.1166	1.000

*-p. < .05 **-p. < .01

("OE" = "open-ended")

Correlation matrix of scale scores
and category variables for closed-ended first survey form

Table 10

	global item	enthus scale	interact scale	learn scale	organ scale	enthus OE	interact OE	learn OE	organ OE
global item	1.0000								
enthusiasm scale	.6975**	1.0000							
interaction scale	.6584**	.7775**	1.0000						
learning scale	.7188**	.6375**	.6110**	1.0000					
organization scale	.7197**	.6175**	.6528**	.7215**	1.0000				
enthusiasm OE	.2325*	.2089*	.0593	.1240	.1530	1.0000			
interaction OE	.3179**	.2591**	.3274**	.2203*	.2582**	-.0130	1.0000		
learning OE	.3326**	.2708**	.3087**	.3181**	.2662**	-.1066	.1399	1.0000	
organization OE	.4223**	.2303*	.2652**	.3803**	.4815**	.1163	.0747	.1059	1.000

*-p. < .05 **-p. < .01

("OE" = "open-ended")

Correlation matrix of scale scores
and category variables for the open-ended first survey form

Table 11

Means and Standard Deviations for

the global item, scale scores and category variables

	mean	**standard deviation**
global item	4.23	.88
enthusiasm scale	4.35	.71
interaction scale	3.43	.99
learning scale	3.88	.95
organization scale	3.98	.97
enthusiasm open-ended	.26	.60
interaction open-ended	.23	.85
learning open-ended	.35	.86
organization open-ended	.20	1.36

The fourTagging Dictionaries used in the content analysis

Each of the following tagging dictionaries were derived from responses to the open-ended items. The same tagging dictionary was used for responses to both the positive and negative open-ended items. The "*" at the end of some of the words tells the software application to look for any word that begins with the respective series of letters. An example would be the word "intimidat*" in the Interaction tagging dictionary. The computer would look for any form of this word (e.g. "intimidate," "indimidating," etc.) in the text file containing the responses.

Table 12

>>ENTHUSIASM<<
enthu*
excite*
enjoy*
gestur*
fun*
humor*
joke
cheer*
passion
exuberance
confidence
added life
forceful
upbeat
want to be here
not just a job
bland
pep
drags on
spice
positive
uplifting

Table 12 (cont.)

>>INTERACTION<<	>>LEARNING<<	>>ORGANIZATION<<	
ignorant	dull	precise	abstract
values the opinions	drowsy	determin*	what is important
individual*	*sleep*	understand*	understood
friend*	awake	to the point	follow*
open	bor*	holistic	integrat*
camarad*	tedious	clear*	thorough*
extra time	exampl*	prepar*	efficient*
level	*interest*	*organiz*	well planned
interact*	informativ*	completely	concise*
extra help	outside material	aids	detail*
approach*	stor*	makes it known	logical
availab*	applic*	easier	direct
personable	retain	flow	illustrat*
down to	additional information	straightforward	explanation*
spends time	clarify	outline*	*relevant
care	stimulat*	structur*	study guide*
pays	practical	relate*	tied in
encourage*	engag*	ties in	tangent*
participat*	thought provoking	reiterat*	distinguish*
ingnorant	varied	clarification	contradict*
bother*	situation*	vague	around the subject
respect	motivat*	on time	drift*
belittle*	learn	syllabus	over my head
feel comfortable	outside information	connect*	punctual
willing	resource*	deadlines	scatter*
contact	experiment*	coherent	figure out
place	demonstr*	changes the schedule	simplified
stupid	analogies	order*	chaotic
beneath	remember*	bounc*	convey her thoughts
ask questions	engross*	changes subjects	define*
repoir	absorb	skips around	handouts
helpful	visual	stray	run over
talk to you	video	time management	over our heads
intimidat*	article*	break it down	lost
name	experienc*	gets off the subject	specific
personally	more materials	written presentation	jumping
feel included	limited	keeps us over	varies
lectures too much	practice	fits	stick
involv*	auditory		
communication	audios		
sarcastic	take more time		
assuming	films		
me thing	tapes		
lectures the whole	data		
students to comment			
continuous lecture			
alternative source			
afraid			